edusolve

# High-Dosage Tutoring:
## Evidence-Based Framework for Scale

KATIE KOSTIN AND NICOLE YARBROUGH, EDUSOLVE
FEBRUARY 2026

# EXECUTIVE SUMMARY AND IMPLICATIONS

A May 2024 survey of public schools by the National Center of Educational Statistics (ies.ed.gov), reported that 46% of public schools in the United States offer high-dosage tutoring (HDT) programs within their district. This marks a 7% increase in such programming from October 2023. To meet the threshold of HDT requirements, programs require consistent and significant funding. Moreover, HDT opportunities can be limited to a single subject area or grade level limiting the significance of the intervention.

The positive effects of well run HDT programs are apparent as measured in numerous academic studies. However, issues with sustainability and program scaling persist in concert with the high cost per pupil to offer HDT. This coupled with current sweeping changes to school funding, including the dismantling of the Department of Education, and the exhaustion of ESSER funds can put impactful HDT programs at risk.

School systems may be required to make cuts if funding is not readily available to sustain additional student support despite their record of success. The redistribution of funds could mean the end of effective HDT interventions impacting students across the United States. In this paper, we provide evidence to support HDT programming as an integral component to effective learning, particularly for students who lack access to private tutoring or individualized support beyond the school day.

This paper evaluates effective tutoring programs through expert analysis of academic publications and thorough examination of empirical data. Readers will be provided with information about the opportunities and barriers to HDT programming and the potential to sustain current offerings, and scale HDT across subject matter and grade level.

# WHY TUTORING?

The benefits of tutoring have long standing acceptance across all fields of learning. Historically access to tutoring was limited to those who could afford the expense of private instruction. The systematic study of tutoring despite economic barriers has yielded promising results and evidence indicates effective tutoring programs can result in significant gains to academic performance for K-12 students. However, multiple barriers to access of quality instruction persist. When a school prepares to invest into a tutoring program they must consider pertinent factors including the curriculum used, selection of tutors, and the optimal setting and dosage for tutoring sessions. The investment of these factors is equivalent to the size of the school district, as school size and resource investment are interdependent.

The purpose of this white paper is to consolidate current scientific data on high-dosage tutoring (HDT) and provide guidance to school and community leaders on best practices in implementation. We aim to provide conclusive evidence on the current state of HDT which can serve to inform best practices and strategic investment to garner the greatest impact for at risk students. Our recommendations are based on evaluation and synthesis of salient academic inquiry.

# WHAT IS HIGH-DOSAGE TUTORING?

For the purpose of this paper, we sought a comprehensive definition of the term, “high-dosage tutoring (HDT).” Through our inquiry, we have evaluated numerous sources that investigated HDT, also referred to as high intensity tutoring. Most studies did not offer a general definition of HDT, but provided definitive parameters for the scope of their investigation that aligned with HDT parameters. After evaluating multiple works, we found Hamlin, (2024) and Kraft et al., (2024), to have thorough HDT criteria. We have integrated definitions from these sources to the following: *HDT requires tutors trained to implement programs aligned with evidence based practices with a 1:1 and a maximum of 4:1 student to tutor ratio. Sessions take place in person during school hours for a minimum of 30 minutes three (3) times per week. HDT programs run for a minimum of 10 weeks during the school year.*

The definition of HDT is the product of decades of study to accurately identify the setting, intervention agents, time and dosage of interventions that produce the greatest effect on academic achievement. It may be true that some academic discourse does not align completely with our definition, however the margin of difference is negligible.

Evidence-based, high-impact interventions that may be utilized in place of, or in addition to high-dosage tutoring were not evaluated in our research. It is necessary to clarify that programs not meeting the criteria above are not included in our research or recommendations. Summer learning programs, after school learning, homework help, extended instructional time (Hamlin, 2024) and other forms of tutoring (e.g., virtual, peer, homework help, test prep) were not included in our definition or inquiry.

| Setting | Curriculum | Tutor | Ratio | Session time | Duration |
|---|---|---|---|---|---|
| In-person, in the school building during the school day | Evidence based aligned to classroom curriculum | Teacher, paraeducator or adult volunteer who has completed training on program implementation. | 1:1 or small groups with a maximum 4:1 ratio | 30-60 minutes | 10 weeks minimum |

# WHAT THE EVIDENCE SAYS

## Method/Search Procedures

The information in this paper reflects a rigorous search of academic literature. Searches were conducted via Google Scholar and ERIC. Search terms used included, “tutoring” AND “high-dosage” OR “high impact” OR “effect.” Searches were limited to work published after 2021. Ancestral searches of select meta-analyses and systematic reviews were conducted for relevant literature published prior to the search parameters. A total of 41 publications were retrieved for screening.

Our search was designed to capture the breadth of current academic evidence of the benefits of high dose tutoring, as well as the barriers to access and success of implementation scaling across large school districts. For this, we examined randomized control trials (RCT) systematic reviews, meta-analyses, and exploratory studies.

20 of the 41 publications were excluded in title and abstract review. We excluded dissertations, and studies conducted outside of the United States, exclusion was extended to publications that did not contain citations, articles that only evaluated peer tutoring and early studies that did not include information about tutoring dosage. Full text analysis identified three (3) duplicate citations and one study that could not be accessed via Google Scholar or academic libraries. After exclusion, nine (9) publications were fully coded to extract all evidence of RCT effects.

After our initial coding and analysis, we found our extraction to be weighted toward interventions serving children in early elementary school. We conducted a secondary search seeking publications that studied high-dosage tutoring (HDT) effects in middle and high school. We were able to identify two more studies that included higher age groups.

**Coding**

We sought multiple sources of publications in this inquiry. Once we established our included studies, we created three categories and coding schema for each type of work. The coding of randomized control trials (7) was conducted using Recall AI to screen content and aggregate narrative coding data. After extraction, we coded 25% of randomized control trials (RCT) publications by hand and compared our results with the Recall AI results. Code agreement between AI and the author was 82%. Meta-analyses (2) were coded by the second author with one study blind-coded by the first author for reliability.

The processes of double coding and blind coding supported the rigor of our findings through inter-rater reliability (IRR). IRR is considered best practice in research and was particularly important in this inquiry to ensure the quality of AI assisted findings. The IRR was determined by dividing the total number of coding agreements by the total number of codes. IRR was established to be 80%. Disagreements were resolved through discussion by the researchers and often due to the narrative nature of our coding procedures.

**Analysis**

When our coding was complete, three categories of inquiry emerged: experimental findings (e.g. randomized control trials), conclusions from meta-analyses and recommendations or other findings within briefs, white papers, quasi experimental studies, exploratory, and conceptual designs. We analyzed the work of field experts and consolidated study results. We provide a general landscape of publications, topics, and findings followed by the opportunities for learners exposed to HDT treatment and barriers to implementation.

Randomized control trials (RCT) are considered the “gold standard” in education and psychology to measure the impact of a treatment or intervention (Kazdin, 2024). High-dosage tutoring (HDT) testing through RCTs are a common method researchers use to measure the elemental impacts of HDT programming on a pool of participants (treatment group) against a control group of students (non-treatment group). Our non-exhaustive search of the literature extracted eight (8) peer reviewed RCTs that measured elements of HDT. Study publication dates range from 1998 through 2024 indicating that HDT impacts have been investigated for decades.

**Analysis, continued**

Our search yielded results weighted toward elementary publications. Four (4) publications including participants in Kindergarten-second grade, one (1) study included participants in grades K-5, one (1) study with middle grade participants and one (1) study of high school participants. Our secondary search found two (2) additional RCTs. One included fourth and fifth grade students, and the other was set in a high school.

Systematic reviews and meta-analyses are designed as a protocol driven exhaustive search for all available literature across multiple academic databases. The studies are coded to extract topical information to answer specific research questions. This process serves the dual purpose of summarizing current literature and identifying gaps where research is still needed (Crowther et al., 2010). In our search for HDT literature we identified two (2) meta-analyses that examined the effects of high dose tutoring across participant grade levels, programs, and at scale.

Publications included in our evaluation, but not meeting the criteria of RCTs or meta-analyses, include policy briefs, conceptual frameworks, white papers, observational, and implementation studies. The multiple designs used in these publications offer commentary on HDT that was not captured in RCT or reviews. These topics include impacts on attendance, mentorship, program design implications and scaling of HDT interventions. We utilized elements of these articles throughout our inquiry, but we did not conduct formal coding procedures due to the diversity of publication content and structure.

**Evaluation and Conclusions from Literature**

Randomized control trial (RCT) studies of high-dosage tutoring (HDT) have investigated a comprehensive range of factors that can predict the success of an HDT program. The formula for successful HDT interventions has been developing since the 1990's. An RCT from 1998 (Rimm-Kaufman et al.) evaluated tutoring effects on 42 at risk kindergarteners. While this study was investigating the specific effect of adult volunteers, other characteristics matched the criteria that would become essential to HDT including 1:1 ratio, 45 minute sessions within the school three (3) times a week for 32 weeks. The results of this design indicated older students who were developmentally ready for the intervention made significant gains; however, younger student outcomes and the limited sample size yielded moderate or no significant effects.

**Evaluation and Conclusions from Literature, continued**

This early evidence may not tout highly significant findings, but it provides a description of the emerging criteria for highly impactful tutoring. Moreover, study results highlight the importance of effective program elements. The inquiry provided insight into age and developmental needs for effective curriculum design to support at risk early learners. The authors highlight that older participants had more significant gains than their younger grade level peers indicating a readiness for the material presented is essential for effective tutoring sessions.

Further inquiry to support HDT development includes the implementation of evidence based practices in the tutor model, and analysis of dosage effects. A three (3) condition RCT using the Tutor Assisted Intensive Learning (TAILS) strategy compared two (2) weekly treatments and four (4) weekly treatments of TAILS intervention to a control group exposed only to storytime (Al Otaiba et al., 2005). Experimental results confirmed that frequency is an essential element in a tutoring program as the participants who had four (4) weekly treatments made significantly greater gains in distal testing measures through the Woodcock Reading Mastery Test than the control and two (2) day a week treatment groups.

Contemporary research has taken the HDT formula and applied it in varying contexts to measure effects. A resurgence of academic inquiry of HDT took place after reports of pandemic related learning loss. School leaders supported by state and federal funding sought out effective interventions to increase student performance (Hamlin, 2024). Scientific inquiry emerged from studies set in schools implementing HDT interventions after the pandemic which broadened the scope of information available to validate numerous HDT outcomes. We have identified studies measuring the effects of tutor implementation, proximal and distal assessment scores, student attendance, optimal dosage and group size, intervention setting, and the impact of specific programs on student learning among other variables. The RCTs evaluated in this paper typically measured multiple factors of HDT impact within the study. Moreover, all studies included “at risk” participants who shared one or more of the following criteria; low performing in academics, attend school in a high poverty area, are eligible for free or reduced lunch, or members of underrepresented communities. Table 2 provides outcomes from the RCTs included in our inquiry.

| Author/Date | Sample Size | Outcomes |
|---|---|---|
| Al Otaiba, S., Schatschneider, C., & Silverman, E. (2005). Tutor-assisted intensive learning strategies in kindergarten: How much is enough?. | 73 Students Kindergarten | -Two dosage conditions compared<br>- 1:1 tutor sessions 4x per week had a significantly greater impact on reading achievement test scores than 1:1 tutor session 2x per week |
| Kortecamp & Peters, (2024). The impact of a high-dosage tutoring program on reading achievement of beginning readers: A multi-level analysis | 185 Student Grades K-1 | -Measured effects of HDT with the Chapter One program<br>-Students assessed on multiple achievement measures and significant improvement was noted in treatment group after one year |
| Jones & Christian (2021). The results of a randomized control trial evaluation of the SPARK literacy program: An innovative approach that pairs one-on-one tutoring with family engagement. | 576 Students Grades K-2 | -Measured effects of HDT with the SPARK literacy program<br>-Students assessed on multiple achievement measures and significant improvement in both literacy skills and attendance was noted in treatment group |
| Rimm Kaufman, S. E., Kagan, J., & Byers, H. (1998). The effectiveness of adult volunteer tutoring | 42 Students Grade 1 | - Tutored children showed greater improvement in letter identification and overall reading compared to controls.<br>- Older tutored children (6.5 years and older) showed more gains than younger ones.<br>- The program was equally effective for English language learners and native speakers. |
| Smith, T. M. et al., (2013). Evaluating math recovery: Assessing the causal impact of a diagnostic tutoring program on student achievement. | 922 Students Grade 1 | - Strong positive effects on the MR diagnostic assessment immediately after tutoring.<br>- Small to moderate positive effects on external standardized math assessments at the end of first grade.<br>- No significant effects on any measures by the end of second grade (fade-out effect). |
| Fryer, & Howard-Noveck (2017). High-Dosage Tutoring and Reading Achievement: Evidence from New York City. | 1,700 students Middle School (6th grade) | -Literacy intervention<br>-Increase in books read, pages read<br>-Increase in student attendance |
| Guryan, et al. (2023). Not Too Late: Improving Academic Outcomes among Adolescents | 5,000+ students Grades- 9-10 | -Improved math achievement test scores.<br>-GPA increase across academic subjects<br>-Decline in class failure, math and other subjects |

**Barriers**

Our search of the literature yielded a notable absence of studies including participants in middle grades and high school. While the results of the randomized control trials (RCT) we evaluated were promising, more literature is needed to establish firm recommendations for these age groups. This is particularly important as high-dosage tutoring (HDT) cannot be defined as an evidence based practice for students in higher grades until the evidence base is established. This will require increased studies on the effects of HDT programming on the middle and high school level.

Inquiry into the effects of HDT should use multiple evaluation measures including both proximal and distal testing. Proximal tests are created to assess skills acquired within a curriculum, and can provide important data about effective instruction. They offer insight into what a student has mastered within a prescribed program. Rigorous studies cannot rely on proximal testing measures alone as they do not demonstrate generalization of skills. Robust studies can include proximal assessment data but must also provide pre- and post- tests using distal, evidence based assessments.

When controlled for other student characteristics, data provided evidence of a positive increase in student attendance at school (Fryer & Howard-Noveck, 2017; Jones & Christian, 2021). However, lack of attendance data or persistent absenteeism during the intervention was noted as a limitation in multiple studies (Al Otaiba et al., 2005; Lee et al., 2024; Williams et al., 2023).

Nickow et al., (2023), use the term “black-boxed” to describe the curriculum used throughout HDT publications. Researchers provide information on study design including dosage, group size, skills addressed and settings followed by intervention outcomes. The limited information about teaching components within a curriculum leaves a gap in our ability to fully understand HDT methods.

Cost may be the most pervasive barrier to sustainable application of HDT. The issue of cost is typically not addressed in experimental design, but is paramount to the practical application of HDT. Cost per pupil is increasingly evident in contemporary publications which addressed program scaling (Nickow et al., 2023; Bhatt et al., 2024; Kraft et al., 2024). Moreover, studies in HDT have been conducted in urban areas where resources are readily available, more research is needed to assess cost of HDT in rural districts (Hamlin 2024).

**Why Results Break at Scale**

The high cost per student of high-dosage tutoring (HDT) programming will not be sustainable for most school districts who have been tackling the changing landscape of district funding. Funding can be in peril due to the lack of empirical evidence supporting scalable programs and sustained student gains. Districts who see the value in HDT programming may respond by seeking ways to lower costs in order to maintain some level of support. Changes to the instruction, dosage, tutor model or other components of a successful HDT model can inadvertently reduce the quality and result in lower measurable impacts (Kraft et al., 2024).

Empirical evidence of HDT is typically measured with randomized control trial (RCT) experiments. RCT participation requires informed consent from both the child and their parents or guardian. When and how parental consent was provided may indicate the level of parental involvement and support for study participants beyond the measurable parameters of the inquiry (Kelpp, 1995). Evidence based information of HDT interventions relies on RCT data. Therefore, when large datasets are analyzed for scaled results of HDT, they include the performance of students who experience varying levels of parental support. This claim is clarified when comparing a study of the SPARK program using RCT in which parent engagement is a measurable component (Jones et al., 2021) to data aggregated from CDC reports comparing the impact of parental behavior on post-pandemic learning loss for children (Hamlin, 2024). In the SPARK study, parents received guidance on in-home literacy engagement strategies indicating a continuum of support across participant environments. Conversely, analysis of 2021 CDC student survey data (Hamlin, 2024) indicated 55 percent of high school students reported experiencing emotional abuse at home. In comparison, these two studies illustrate the polarity of parental support that students experience which may explain some discrepancies in the data for HDT at scale.

With the current literature available on scaling, there is not a consensus on why programs break at scale or how to improve results when programs expand. However, based on the evidence we can suggest that district leaders focus on implementation quality and not bottom line cost. We also propose scaling to include parent engagement elements. Finally district leaders should include statistical controls for effects of the continuum support provided across student environments and other factors that may impact outcomes of program scaling beyond large scale assessment data.

**Provider Comparison Matrix**

Industry and government organizations including What Works Clearinghouse, Cognia, and the National Student Support Accelerator, offer trusted benchmarks for identifying high-quality tutoring interventions. We use these to evaluate Math Recovery (Integrow Numeracy Solutions), Chapter One, SPARK Literacy (Future Forward), Amplify Reading, and Saga Education. These providers were selected because they were reviewed by the academic publications in this paper. We also include One on One Learning in the matrix because its program design aligns with the high-dosage tutoring (HDT) principles defined in this paper and because it is a vendor used by the focal district.

The Every Student Succeeds Act (ESSA) of 2015 requires states, districts, and schools to provide instruction and student support with evidence-based, research-grounded programs. ESSA requirements must be satisfied for schools to receive Title funds. Eligibility for funding is measured by student outcome reporting. The main goal is to ensure equitable, high-quality education for all students. By selecting proven interventions, K-12 leaders increase the probability of student achievement gains. Five factors determine ESSA evidence tiers: study design, results, findings from related studies, sample size and setting, and match. Based on these criteria, programs are classified into one of four tiers: strong (Tier I), moderate (Tier II), promising (Tier III), or demonstrating a rationale (Tier IV).

The What Works Clearinghouse (WWC), operated by the Institute of Education Sciences (IES), supports the ESSA process by reviewing and rating student interventions against rigorous, methodological standards. WWC designations such as Meets Standards Without Reservations and Meets Standards With Reservations often correspond to Tier I and Tier II evidence levels and provide districts with an independent, publicly accessible validation of research quality searchable in their database.

**Provider Comparison Matrix, continued**

Cognia is a nonprofit that supports school improvement through accreditation and certification. It conducts external reviews of K-12 public, charter, early learning, and postsecondary institutions against research-based performance standards. In 2024, Cognia launched tools to evaluate education services, organizations, and extended learning programs using criteria including; culture of learning, leadership for learning, engagement of learning, and growth in learning. Selecting a provider that meets Cognia's criteria or pursuing review for a site strengthens confidence that the tutoring program is a recognized, effective intervention.

The National Student Support Accelerator (NSSA) provides tools, research, and technical assistance for tutoring organizations and K-12 systems to scale "high-impact" tutoring. Our HDT parameters align with NSSA's core principles but establish specific, evidence-based minimum thresholds drawn from research in our academic paper reviews, while NSSA's approach remains broader to test variations. NSSA awards "Program Design Badges" to providers who fully meet high-impact standards, offering districts an additional indication of quality alignment.

| Provider | ESSA Tier | WWC Status | NSSA Status | External Quality Validation |
|---|---|---|---|---|
| One on One Learning | Tier II (Strong): Quasi-experimental evidence (vendor cited) | No formal WWC rating; aligns with WWC guidance | High Impact Tutoring (HIT) Program Design Badge | Cognia accredited |
| Saga Education | Tier I (Strong): 3 RCTs, avg. effect +0.14 | None | HIT 6-12 Math Program Design Badge | None |
| Chapter One | Tier I (Strong): RCTs show +0.23 - 0.31 literacy gains | None | HIT Program Design Badge | None |
| SPARK Literacy (now Future Forward (FF)) | Tier I (Strong): Multiple RCTs for reading and attendance | Foundational SPARK study meets WWC Standards Without Reservations (IES 32028) | Future Forward Literacy Program Design Badge | None |
| Math Recovery (now Integrow Numeracy Solutions (INS)) | Tier II (Moderate): Evidence through PD intervention studies | No formal WWC rating; aligns with WWC guidance | None | None |
| Amplify (reading program now Boost Reading) | Tier I (Strong): 1 RCT; additional quasi-experimental evidence (vendor cited) | None | Amplify Tutoring Program Design Badge | None |

**Provider Comparison Matrix, continued**

Of the six providers reviewed, One on One Learning is the only organization holding Cognia accreditation, representing external validation of organizational quality, implementation systems, and instructional supports beyond research evidence tiers.

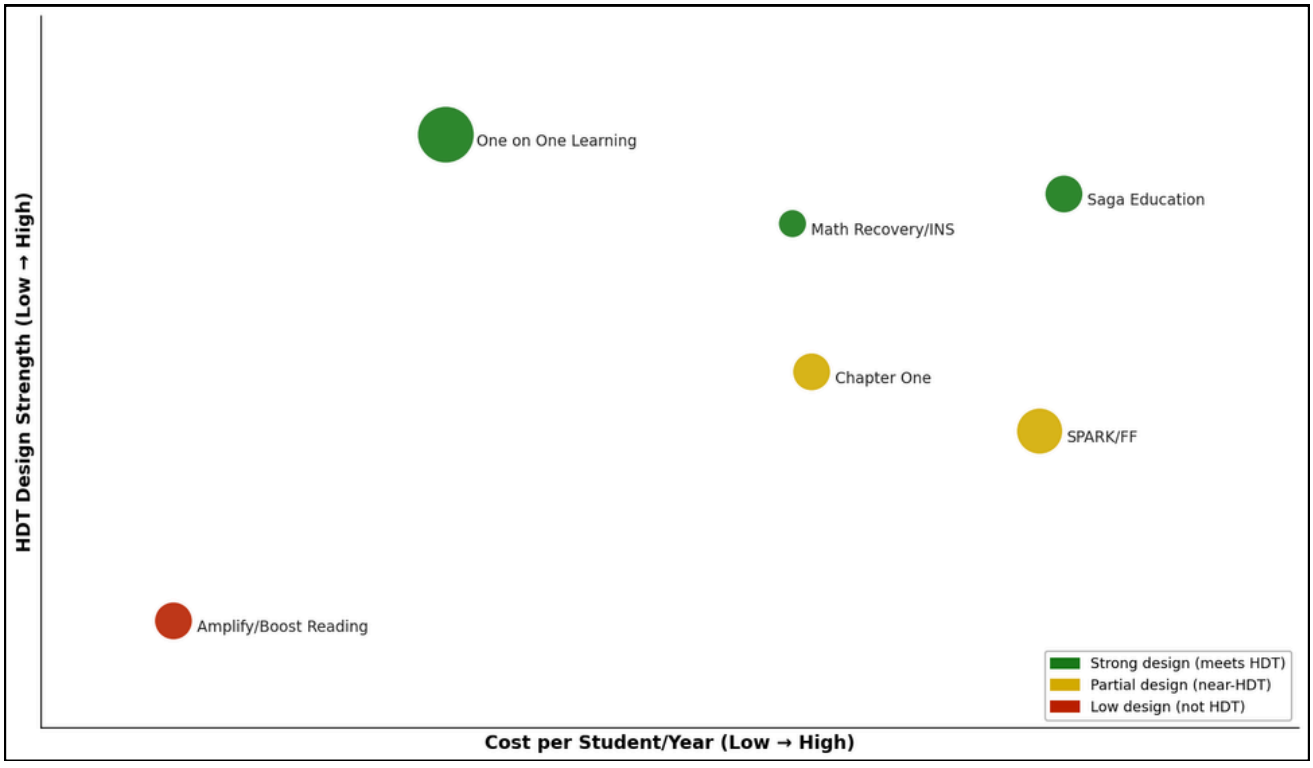
A tutoring program is considered evidence-based when its model is supported by research that meets ESSA tiers, validated by WWC standards, or is recognized by NSSA's design criteria. The table below applies the operational definition of high-dosage tutoring established earlier in this paper to a comparison matrix. It evaluates each program's alignment with the research-based criteria essential for the efficacy of a provider. Programs vary substantially in cost and in the degree to how closely their designs align to HDT research. They also differ in the strength of evidence supporting their efficacy.

| Provider | Evidence-Based/ Validated | Trained Tutors | Ratio ≤ 4:1 | In-school and in-person | ≥ 30 min | ≥ 3x p/week | ≥ 10 weeks | Meets HDT |
|---|---|---|---|---|---|---|---|---|
| One on One Learning | ESSA+WWC+NSSA + Cognia | ✅ | ✅ | ✅ | ✅ | ✅ | ✅ | YES |
| Saga Education | ESSA + NSSA | ✅ | ✅ | ✅ | ✅ | ✅ | ✅ | YES |
| Math Recovery/ INS | ESSA + WWC | ✅ | ✅ | ✅ | ✅ | ✅ | ✅ | YES |
| Chapter One | ESSA+ NSSA | ✅ | ✅ | ✅ | ⚠️ | ✅ | ✅ | PARTIAL |
| SPARK Literacy/ FF | ESSA+WWC+NSSA | ✅ | ⚠️ | ✅ | ✅ | ✅ | ✅ | PARTIAL |
| Amplify (reading program now Boost Reading) | ESSA + NSSA | ❌ | N/A | ⚠️ | ⚠️ | ⚠️ | ✅ | NO |

**Cost, Design Quality, and Evidence Strength Across Tutoring Providers**

The providers included in our primary evidence review were identified through academic publications. The comparison matrix includes One on One Learning, which was not part of the academic sample but was added because it aligns with HDT design principles and is a vendor by the focal district. Its inclusion illustrates how the proposed rubric can be applied to vendors that meet HDT benchmarks even when formal impact evaluations are not yet available in the peer reviewed literature.

In practice, districts must choose among many vendors, including those without strong published evidence. By applying consistent evidence and design benchmarks to both programs with published impact studies and to vendors like One on One Learning whose models align with HDT principles, districts can more reliably prioritize tutoring approaches that are most likely to meet HDT standards.




- Bubble size indicates number of validated evidence sources

# IMPLICATIONS FOR DISTRICTS

High-dosage tutoring (HDT) is an evidence based practice (EBP). Like all EBPs, it provides promising, evidence supported models to improve academic performance for students defined as at risk and in need of high impact academic intervention. HDT implementation can vary across school districts and student populations. School leaders must leverage the needs within their student population to the budgetary restraints and supplementary support available within their district and community. To support this effort, the the tools and guidance listed below can support the commencement, improvement, and scaling of district HDT programming.

- School leaders seeking to continue supporting their at-risk population despite decreased funding can seek guidance from the National Student Support Accelerator (NSSA). NSSA is providing school leaders guidance through open access materials, current research, self assessment tools, and strategic guidance on the design and implementation of district wide HDT.
- Continued implementation and scaling of cost effective HDT is achievable with strategic innovation. Literature on the effects of HDT emphasizes the need for in-person tutoring models and does not account for the scope of technological advancement. However, the Saga High Impact Math Tutoring with a computer assisted learning platform, has shown positive effects on high school students with decreased in person tutoring (Bhatt et al., 2024). School leaders can benefit from increased awareness of technological advancement and integration into HDT design.
- School districts can benefit from community partnerships that focus on HDT implementation. Community based organizations (CBOs) who partner with school districts can use their platform to advocate for HDT initiatives and identify funding sources (Bronson & Krajewski, 2025). Moreover CBOs can provide a low cost solution for trained in-person tutors to deliver HDT.
- When building and scaling HDT within a district, leadership will need to pinpoint evidence based, cost effective programming that is the best fit for their student population. Rather than make an all in, long term investment districts can consider a pilot program strategy to compare the effects of multiple programs. The Strategic Data Project provides tools and resources for program assessment and data sharing within or across districts to assess what works and where improvement is required.

# APPENDIX

## Procurement Rubric

We developed a comprehensive scoring rubric to provide a way for procurement evaluators to compare tutoring programs on the features that most strongly predict impact, rather than based on marketing claims or cost alone. The rubric guides evaluators to distinguish between programs that merely “offer tutoring” and those that meet the research-based definition of high-dosage tutoring (HDT). Using this table as a model supports transparent, consistent decisions across multiple providers and aligns procurement selection with the evidence base summarized in this paper.

## HDT Supplemental Feedback Checklist

The following checklist can serve as a useful supplemental feedback tool for teachers and school-based staff to use during the procurement or RFP process as they reflect on how a proposed tutoring program will function in their specific context. It foregrounds practical implementation features that the procurement rubric cannot fully capture, such as whether the program supports flexible scheduling during or after the school day, covers the instruction levels present in the building (early learning, elementary, middle, high school), and offers reproducible or openly available lesson materials that reduce ongoing costs and preparation time. In addition, the sections on activity elements and cultural understanding invite educators to consider whether embedded practices are evidence-based, aligned to their school curriculum, provide user-friendly progress monitoring interoperability, representation, student interests, and a warm, welcoming learning environment. Used alongside the weighted scoring rubric, this yes/no checklist can help leaders and decision makers surface school-level insights about feasibility, fit, and cultural responsiveness that should inform procurement processes.

## Annotated Summary of Key Sources

The Annotated Summary of Key Sources provides brief summaries of the most important studies and sources used to support the definition, criteria, and recommendations for high-dosage tutoring (HDT) presented in this white paper. Each entry includes the full citation in APA 7th edition format, a summary of the main findings, and an explanation of how the source informed our analysis.

# PROCUREMENT RUBRIC

## Procurement Rubric

This comprehensive scoring rubric provides a way for procurement evaluators to compare tutoring programs on the features that most strongly predict impact. The rubric guides evaluators to distinguish between programs that merely "offer tutoring" and those that meet the research-based definition of high-dosage tutoring (HDT). Using this table as a model supports transparent, consistent decisions across multiple providers and aligns procurement selection with the evidence base summarized in this paper.

| SCORING INTERPRETATION | |
|---|---|
| **3.20-4.00 (80-100%)** | Strong Candidate - Exemplary HDT alignment with Tier I evidence |
| **2.40-3.19 (60-79%)** | Acceptable Candidate - Meets most HDT criteria with solid evidence |
| **1.60-2.39 (40-59%)** | Developing - Partial alignment; significant gaps in design or evidence |
| **<1.60 (<40%)** | Not Recommended - Does not meet HDT definition; reconsider classification |

## Weight Rationale

Weights reflect the relative importance of each dimension based on synthesized evidence from randomized control trials (RCT), meta-analyses, and implementation studies reviewed in this white paper. Dosage & Frequency receive the highest weight (20%) due to their consistently strong association with effect sizes. Core implementation elements (Tutor Type, Training, Curriculum, Data Use) are equally weighted (10% each) as they interact to determine quality. Evidence Strength (ESSA/WWC/NSSA) validates external credibility but is downstream of design fidelity (10%). Model, Group Size, Duration, and Cost receive lower weights (5-10%) as threshold criteria.

# PROCUREMENT RUBRIC

| Criterion | Weight | 1 - Not Met | 2 - Partially Met | 3 - Meets | 4 - Exceeds | Score (1-4) | Weighted |
|---|---|---|---|---|---|---|---|
| **Model (Setting)** | 10% | Virtual-only implementation; no in-school option | After-school only OR primarily virtual with optional school-day access | In-person implementation during the school day in most sites | Systematic in-school implementation with strong integration into master schedule across sites | | 0 |
| **Tutor Type** | 10% | Predominantly untrained volunteers or peer tutors with minimal oversight | Mix of paraprofessionals or non-certified adults with some training | Teachers or paraprofessionals with defined selection criteria and training | Certified teachers or experienced educators, selected and supervised with clear quality standards | | 0 |
| **Group Size (Ratio)** | 10% | Typical group size >4:1 | Group size sometimes at or below 4:1 but frequently higher | Consistent 1:1 to 1:4 groups with documented expectations | Predominantly 1:1 to 1:2 with strong fidelity monitoring to keep ratios low | | 0 |
| **Dosage & Frequency** | 20% | <60 min/week OR ≤1 session/week | 60-89 min/week AND 2 sessions/week | ≥90 min/week across ≥3 sessions per week (e.g., 3×30 min) | ≥120 min/week AND ≥4 sessions/week with documented fidelity monitoring | | 0 |
| **Duration** | 5% | Fewer than 8 weeks | 8-9 weeks total | 10-20 weeks within the school year | Full semester or full-year implementation for most participants | | 0 |
| **Training Hours** | 10% | <5 hours of initial training; no ongoing support | 5-19 hours of initial training OR limited follow-up | 20-39 hours of initial training with scheduled coaching or PLCs | ≥40 hours of structured initial training PLUS regular coaching and observation | | 0 |

# PROCUREMENT RUBRIC

| Criterion | Weight | 1 - Not Met | 2 - Partially Met | 3 - Meets | 4 - Exceeds | Score (1-4) | Weighted |
|---|---|---|---|---|---|---|---|
| **Curriculum Alignment** | 10% | No clear alignment to state standards or core curriculum | General standards alignment, but weak linkage to district core materials | Explicit alignment to grade-level standards and core curriculum sequences | Alignment PLUS embedded supports (shared texts, co-planning tools, integrated assessments) | | 0 |
| **Data Use & Progress Monitoring** | 10% | No systematic progress monitoring | Periodic or teacher-initiated checks only | Regular progress monitoring (every 2-4 weeks) with guidance for instructional decisions | Integrated formative assessment and dashboards with adaptive recommendations and district data integration | | 0 |
| **Evidence Strength (ESSA/WWC /NSSA)** | 10% | No ESSA Tier I-III study; WWC 'Does Not Meet Standards' or not reviewed; No NSSA badge | At least one promising/emerging study (ESSA Tier III) OR vendor quasi-experimental; WWC not reviewed or limited; No NSSA badge but listed | At least one ESSA Tier II (or multiple Tier III) study OR WWC 'Meets With Reservations'; Clear alignment to NSSA principles | Multiple ESSA Tier I RCTs AND/OR WWC 'Meets Without Reservations' PLUS NSSA Program Design Badge | | 0 |
| **Cost per Student** | 5% | >$2,500 per student per year | $1,501-$2,500 per student per year | $501-$1,500 per student per year | ≤$500 per student per year for a full HDT model | | 0 |
| **TOTAL SCORE** | **100%** | | | | | | **0** |

# HDT SUPPLEMENTAL FEEDBACK CHECKLIST

The following checklist can serve as a useful supplemental feedback tool for teachers and school-based staff to use during the procurement or RFP process as they reflect on how a proposed tutoring program will function in their specific context. It foregrounds practical implementation features that the procurement rubric cannot fully capture, such as whether the program supports flexible scheduling during or after the school day, covers the instruction levels present in the building (early learning, elementary, middle, high school), and offers reproducible or openly available lesson materials that reduce ongoing costs and preparation time. In addition, the sections on activity elements and cultural understanding invite educators to consider whether embedded practices are evidence-based, aligned to their school curriculum, provide user-friendly progress monitoring interoperability, representation, student interests, and a warm, welcoming learning environment. Used alongside the weighted scoring rubric, this yes/no checklist can help leaders and decision makers surface school-level insights about feasibility, fit, and cultural responsiveness that should inform procurement processes.

| Criterion | Supplemental Feedback Checklist | Rating |
|---|---|---|
| **Program Elements** | Small group 1:1 to 1:3 | Y. N. NA |
| | Flexible implementation- during or after school hours | Y. N. NA |
| | In person implementation format | Y. N. NA |
| | Design allows for 3 lessons per week | Y. N. NA |
| | Instructional level | |
| | Early learning | Y. N. NA |
| | Elementary | Y. N. NA |
| | Middle Grades | Y. N. NA |
| | High School | Y. N. NA |
| | Lessons provide 30-60 minutes of instruction | Y. N. NA |

# HDT SUPPLEMENTAL FEEDBACK CHECKLIST

| Criterion | Supplemental Feedback Checklist | Rating |
|---|---|---|
| **Activity Elements** | Embedded evidence based practices | Y. N. NA |
| | Evidence of alignment to school curriculum | Y. N. NA |
| | Offer progress monitoring within activity matrix | Y. N. NA |
| | Built in intervention and scaffolding for learning targets | Y. N. NA |
| | Evidence of student progress on distal assessments | Y. N. NA |
| | Lesson materials are reproducible or openly available | Y. N. NA |
| **Tutor procurement and training** | Tutors require an advanced degree in teaching | Y. N. NA |
| | Tutor model is designed for any school staff (e.g. teachers, paraprofessionals) | Y. N. NA |
| | Tutor model can include volunteers or nonparental adults | Y. N. NA |
| | Program provides a peer tutor model | Y. N. NA |
| | Training is free | Y. N. NA |
| | Training is fee based | Y. N. NA |
| | Program provides tutor support via email or chat | Y. N. NA |
| | Training completed online | Y. N. NA |
| | Training in person | Y. N. NA |
| **Cultural understanding** | Includes high interest material for all learners | Y. N. NA |
| | Representation and diversity are demonstrated in content and strategies | Y. N. NA |
| | Individualization of content and strategies | Y. N. NA |
| | Tutor training emphasizes a warm and welcoming environment | Y. N. NA |
| | Flexible content inclusive of student interests and learning style | Y. N. NA |

# ANNOTATED SUMMARY OF KEY SOURCES

**Meta-Analyses and Systematic Reviews**

Nickow, A. J., Oreopoulos, P., & Quan, V. (2024). The promise of tutoring for PreK–12 learning: A systematic review and meta-analysis of the experimental evidence. American Educational Research Journal, 61(1), 74–107.

***Summary.*** Comprehensive meta-analysis of 96 randomized controlled trials examining tutoring interventions across PreK–12 settings. Found an average effect size of 0.37 standard deviations, with stronger effects for programs featuring small group sizes (1:1 to 1:3), frequent sessions (≥3 times per week), trained tutors, and alignment with classroom curriculum.

***Relevance.*** This meta-analysis provided the empirical foundation for our HDT definition, particularly the emphasis on dosage, frequency, and group size. The study's analysis of moderator variables informed our weight allocation in the procurement rubric, with dosage and frequency receiving the highest weight (20%) based on their strong association with effect sizes.

Kraft, M. A. (2024). What impacts should we expect from tutoring at scale? Tutoring Meta-Analysis [EdWorkingPaper]. Annenberg Institute at Brown University.

***Summary.*** Meta-analytic synthesis examining how tutoring effectiveness varies when programs are scaled beyond controlled research settings. Documents the phenomenon of "results breaking at scale" due to implementation challenges, tutor quality variation, and reduced fidelity to evidence-based design principles.

***Relevance.*** Used to frame the "Why Results Break at Scale" section of the white paper and to inform our discussion of implementation support, fidelity monitoring, and the tension between research efficacy and real-world effectiveness. Influenced our emphasis on evidence strength (ESSA/WWC/NSSA) as a procurement criterion.

**Randomized Controlled Trials**

Al Otaiba, S., Schatschneider, C., & Silverman, E. (2005). Tutor-assisted intensive learning strategies in kindergarten: How much is enough? Exceptionality, 13(4), 195–208.

***Summary.*** RCT with 73 kindergarten students comparing two dosage conditions: 1:1 tutoring 4 times per week versus 2 times per week. Students receiving 4 sessions per week made significantly greater gains in reading achievement test scores than those receiving 2 sessions per week. Control group received small-group storybook reading.

***Relevance.*** Critical evidence supporting the importance of frequency in HDT programming. Directly informed our operational definition requirement of ≥3 sessions per week and the high weight assigned to dosage and frequency (20%) in the procurement matrix. Demonstrates that insufficient dosage (2 days/week) yields limited impact.

Kortecamp, K., & Peters, M. L. (2024). The impact of a high-dosage tutoring program on reading achievement of beginning readers: A multi-level analysis. Journal of Education for Students Placed at Risk, 29(3), 291–309.

***Summary.*** RCT evaluating the Chapter One program with 185 students across 13 classrooms in an Ohio Title I school. Students receiving 1:1 tutoring 2-3 times per week for brief sessions (5-6 minute conferences plus 10-15 minutes daily independent practice) during the school year showed statistically significant gains across multiple reading measures (iReady, Chapter One RFS assessments, Chapter One ORF, and Fountas & Pinnell) compared to controls receiving only teacher instruction. Treatment students outperformed control students on all end-of-year kindergarten assessments and middle-of-year first grade assessments. Effect sizes ranged from 0.44 to 0.73 depending on the measure.

**Randomized Controlled Trials, continued**

***Relevance.*** Provided evidence for Chapter One as a program meeting our HDT criteria with structured, evidence-based curriculum aligned with classroom instruction and delivered by trained paraprofessionals (college graduates with specialized training). Demonstrated that brief, frequent tutoring sessions combined with independent practice can produce measurable outcomes for at-risk early readers. The study's use of nationally validated assessments (iReady meets ESSA evidence-based criteria) and rigorous RCT methodology informed our emphasis on evidence strength in the procurement rubric. Particularly relevant for demonstrating effectiveness of paraprofessional tutors when properly trained and supported with high-quality curriculum.

Jones, C. J., & Christian, M. (2021). The results of a randomized control trial evaluation of the SPARK literacy program: An innovative approach that pairs one-on-one tutoring with family engagement. Journal of Education for Students Placed at Risk, 26(3), 185–209.

***Summary.*** Two-year RCT with 576 K–2 students in Milwaukee Public Schools. SPARK program combined 1:1 in-school tutoring (up to 3 times per week, 30 minutes per session) with family engagement support. Resulted in significant improvements in reading achievement (effect size 0.23), literacy skills (effect size 0.35), and school attendance (5.8 fewer absences).

***Relevance.*** Demonstrated the dual benefits of HDT for both academic outcomes and attendance. Informed our discussion of attendance as a measurable outcome and the value of family engagement components. SPARK's WWC "Meets Standards Without Reservations" rating and NSSA Program Design Badge made it a key example in our evidence strength criterion.

**Randomized Controlled Trials, continued**

Rimm-Kaufman, S. E., Kagan, J., & Byers, H. (1998). The effectiveness of adult volunteer tutoring. Literacy Research and Instruction, 38(2), 143–152.

***Summary***. Early RCT with 42 first-grade students comparing 1:1 volunteer tutoring (45 minutes, 3 times per week for 32 weeks) to a control group. Tutored children showed greater improvement in letter identification and overall reading. Older children (≥6.5 years) showed more gains than younger peers, indicating developmental readiness matters.

***Relevance.*** One of the earliest studies demonstrating the emerging criteria for effective tutoring dosage and frequency. Highlighted the importance of developmental appropriateness in curriculum design and tutor training. Informed our discussion of tutor qualifications and the historical development of HDT as an evidence-based practice.

Smith, T. M., Cobb, P., Farran, D. C., Cordray, D. S., & Munter, C. (2013). Evaluating Math Recovery: Assessing the causal impact of a diagnostic tutoring program on student achievement. American Educational Research Journal, 50(2), 397–428.

***Summary.*** Two-year RCT with 922 first-grade students across 20 schools. Math Recovery (1:1, ~25 minutes, 4–5 sessions per week, 12 weeks per cycle) showed strong immediate effects on proximal assessments but small to moderate effects on external standardized tests at the end of first grade. Effects faded by the end of second grade.

***Relevance.*** Important for understanding fade-out effects and the limitations of even well-designed HDT programs when ongoing support ends. Informed our discussion of barriers, sustainability challenges, and the need for long-term academic support. Used as an example of Tier II (ESSA Moderate Evidence) program in provider comparison matrix.

**Randomized Controlled Trials, continued**

Fryer, R. G., Jr., & Howard-Noveck, M. (2017). High-dosage tutoring and reading achievement: Evidence from New York City (Working Paper No. 23792). National Bureau of Economic Research.

***Summary.*** School-level RCT with approximately 1,700 middle school (6th grade) students across 60 NYC public schools. Small-group tutoring (4 students per tutor, 45–60 minutes, average 67 days per year over 3 years) significantly increased school attendance (1.2 percentage points) and had positive effects on ELA achievement, particularly for Black students (0.09 SD gain).

***Relevance.*** One of the few rigorous studies of HDT in middle school settings, filling a critical gap in the literature. Informed our discussion of grade-level variation in HDT effectiveness and the importance of attendance as both a mechanism and an outcome. Highlighted differential impacts by student subgroup, informing our equity and accessibility criterion in the procurement rubric.

Guryan, J., Ludwig, J., Bhatt, M. P., Cook, P. J., Davis, J. M. V., Dodge, K., Fryer, R. G., Jr., Levitt, S. D., Mayer, S. E., Pollack, H. A., Steinberg, L., & Stoddard, G. (2023). Not too late: Improving academic outcomes among adolescents. American Economic Review, 113(3), 738–765.

***Summary.*** Two RCTs with over 5,000 high school students (grades 9–10) in Chicago Public Schools. SAGA Education's daily, intensive math tutoring (one class period per day, small groups) significantly improved math test scores (0.19–0.37 SD), math GPA (0.56 points), and reduced math course failures by ~49%. No effects on reading or social-emotional measures.

***Relevance.*** Critical evidence that HDT can effectively support adolescents who are behind academically, challenging the "too late" narrative. Informed our discussion of HDT effectiveness in high school settings and the importance of alignment with students' current coursework. Used to justify high weight on dosage and frequency in the procurement matrix.

**Implementation and Policy Studies**

Kraft, M. A., & Falken, G. T. (2021). A blueprint for scaling tutoring across public schools. (EdWorkingPaper No. 20-335). Annenberg Institute at Brown University.

***Summary.*** Thought experiment exploring how tutoring could be scaled nationally to address COVID-19 learning loss and become a permanent feature of the U.S. public education system. Outlines ten core design principles including: tutoring as school-wide program, individualized instruction (4:1 ratios, preferably 2:1), high-dosage intervention (3-5 times per week, 30 minutes), continuity with same tutor, integration into school day, intensive ongoing training, and curriculum-based approach. Proposes tiered structure using cross-age peer tutors (high school students tutoring elementary, college students tutoring middle school via Federal Work-Study, college graduates tutoring high school via AmeriCorps). Estimates costs of $5-15 billion annually for targeted approaches focusing on K-8 Title I schools or low-proficiency schools.

***Relevance.*** Provided a comprehensive framework for understanding implementation challenges and design tradeoffs when scaling tutoring. Informed our discussion of barriers to scaling, including tutor recruitment, training infrastructure, scheduling complexities, and cost considerations. The emphasis on evidence-based design principles (school-wide programs, low ratios, high frequency, same tutor relationships, school-day integration, intensive training, curriculum alignment) validated and strengthened our operational HDT definition. Cost estimates contextualized our discussion of sustainability and informed the 5% weight assigned to cost per student in the procurement rubric. The analysis of past failures (America Reads, Supplemental Education Services) shaped our recommendations for strong implementation support and fidelity monitoring.

**Implementation and Policy Studies, continued**

Hamlin, D. (2024). From crisis to opportunity: Post-pandemic academic growth in Massachusetts (White Paper No. 276). Pioneer Institute for Public Policy Research.

***Summary.*** White paper synthesizing Massachusetts state data on post-pandemic learning recovery, including analysis of HDT and summer learning programs funded by ESSER. Found that high-dosage tutoring yielded approximately one year of learning gains (0.37 SD) but faced sustainability challenges due to high costs ($3,500–$4,300 per student per year) and tutor recruitment difficulties.

***Relevance.*** Provided current context for HDT implementation challenges post-pandemic, including cost barriers, scaling difficulties, and the impending exhaustion of ESSER funds. Informed our operational HDT definition and discussion of sustainability and cost considerations in the procurement rubric (5% weight for cost per student).

Lee, M. G., Loeb, S., & Robinson, C. D. (2024). Effects of high-impact tutoring on student attendance: Evidence from the OSSE HIT Initiative in the District of Columbia (EdWorkingPaper No. 24-1107). Annenberg Institute for School Reform at Brown University.

***Summary***. Quasi-experimental study using within-student fixed effects with 4,222 tutored students across 141 Washington, DC schools. Scheduled tutoring sessions reduced daily absenteeism by 1.2 percentage points (7% overall reduction), with stronger effects for middle school students and those with high prior absenteeism. In-school delivery and smaller tutor-to-student ratios amplified benefits.

***Relevance.*** Demonstrated that HDT has attendance benefits beyond academic outcomes, supporting its use as a dual-benefit intervention for engagement and learning. Informed our discussion of attendance as a measurable outcome and the importance of in-school implementation (included as a 10% criterion in the procurement matrix).

**Implementation and Policy Studies, continued**

Casalaspi, D., Mission, M., & Korman, H. T. (2025). From policy to impact: Illinois' implementation of high-impact tutoring (Case Study 2). Bellwether.

***Summary.*** Implementation case study of the Illinois Tutoring Initiative (ITI), which served over 7,200 students in grades 3–8 across 209 schools during 2022–2024. Nearly 90% of students met or exceeded expected math growth and 80% did so in reading. Key success factors included strong cross-agency leadership, centralized data management, embedded evaluation, and balance between evidence-based guardrails and local flexibility.

***Relevance.*** Real-world example of successful large-scale HDT implementation. Informed our "Implications for Districts" section and recommendations for strong governance, data use, and continuous improvement. Demonstrated that scaling is achievable with strategic planning and cross-agency coordination.

Pilchen, A., Klute, M., Mielicki, M., & Kumbroch, H. (2024). Examining implementation and outcomes of the Project On-Track high-dosage literacy tutoring program (REL 2024-005). Regional Educational Laboratory Appalachia.

***Summary***. Observational study of 1,126 students (grades 1–3) in northeastern Tennessee receiving Amplify Reading/mCLASS tutoring (1:3 groups, 30 minutes, ≥2 times per week). Among the 622 students at highest risk, 42% improved their literacy risk category after one year. No significant differences in outcomes were found based on tutoring timing (during vs. outside school hours), frequency, or tutor qualifications, suggesting high-quality structured curriculum may allow implementation flexibility.

***Relevance.*** Provided evidence that structured, evidence-based curriculum can allow flexibility in implementation without compromising outcomes. Informed our discussion of curriculum alignment and training hours as separate but related criteria in the procurement rubric.

**Evidence Standards and Frameworks**

What Works Clearinghouse. (n.d.). Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/wwc/

***Summary.*** Independent review body that evaluates education interventions against rigorous methodological standards. Ratings include "Meets Standards Without Reservations" and "Meets Standards With Reservations," corresponding to Tier I and Tier II ESSA evidence levels.

***Relevance.*** Used as one of three pillars (alongside ESSA and NSSA) in our Evidence Strength criterion in the procurement rubric. WWC ratings provide independent validation of research quality and help districts identify programs with strong causal evidence.

National Student Support Accelerator. (n.d.). Stanford University. https://nssa.stanford.edu/

***Summary.*** Organization providing research, tools, and technical assistance for scaling high-impact tutoring. Awards Program Design Badges to programs meeting high-impact standards based on tutor quality, frequency, personalization, and curriculum alignment.

***Relevance.*** NSSA's framework aligned closely with our operational HDT definition but with broader parameters to test implementation variations. NSSA Program Design Badges were integrated into our Evidence Strength criterion and used to validate provider quality in the comparison matrix.

**Methodological References**

Kazdin, A. E. (2024)**.** Randomized controlled trials: Characteristics, options, and challenges.

***Summary.*** Overview of RCT methodology, emphasizing that RCTs are the "gold standard" for measuring intervention impact in education and psychology due to rigorous control of confounding variables.

***Relevance.*** Cited to justify our focus on RCTs as the primary source of causal evidence for HDT effectiveness.

**Methodological References, continued**

Crowther, M., Lim, W., & Crowther, M. A. (2010). Systematic review and meta-analysis methodology. Blood, The Journal of the American Society of Hematology, 116(17), 3140–3146.

***Summary.*** Describes the protocol-driven process of systematic reviews and meta-analyses, which serve to summarize current literature and identify research gaps.

***Relevance.*** Cited to explain the role and rigor of meta-analyses in synthesizing HDT evidence and justifying our reliance on Nickow et al. (2024) and Kraft (2024) as anchor sources.

Klepp, K. I. (1995). Nonresponse bias due to consent procedures in school-based, health-related research. Scandinavian Journal of Social Medicine, 23(1), 53–59.

***Summary.*** Discusses how RCT participant consent procedures may introduce selection bias, as families who consent may have higher levels of parental involvement and support compared to non-consenting families.

***Relevance.*** Informed our discussion of "Why Results Break at Scale," particularly the potential for RCT findings to overestimate effects when scaled to populations with varying levels of parental support and engagement.